CODERPAD BENCHMARK AI · OVERVIEW

Grade what *actually matters.*

Modern technical screens and interviews are open-ended, project-based, and AI-enabled — but pass/fail scoring hasn't caught up.
Benchmark AI bridges the gap with detailed AI insights on the hard-to-measure skills that matter.

As project-based assessments become more complex, **unit tests aren't enough**. Criteria like systems thinking, decision-making, performance, readability, and codebase fit need judgment that pass/fail can't capture. Until now, that meant hours of manual review per candidate. Benchmark AI evaluates this automatically.

WHAT BENCHMARK AI DOES

Compare candidates side-by-side
See how candidates stack up across technical skills, problem-solving, code quality, and communication.

Surface key differentiators
Standout coding approaches and problem-solving patterns — what separates top performers, visible at a glance.

Evaluates how candidates used AI
See which models candidates used, how many prompts they sent, and whether they used AI strategically.

Works at every stage
AI evaluation scores candidates across each stage of the hiring process, giving you a full candidate view from screen to hire.

Evaluates your questions too
Analyzes response patterns across 100s of interviews to show which questions drive the strongest signal.

HOW BENCHMARK AI WORKS · PRODUCT WALKTHROUGH

01
Candidate Summary
Candidate results are segmented into 3 buckets — **Strong, Mixed, and Concerns** — making evaluation easy and giving hiring teams an instant read on pipeline health.

CANDIDATE SUMMARY · API RATE LIMITER · 40 ANALYZED — 3-BUCKET SEGMENTATION

STRONG	MIXED	CONCERNS
15	20	5
38% cleared P3 or P4	50% partial P3 or P4	13% did not complete P2

02
Comparative Evaluation
Every candidate ranked side-by-side with their **results, AI usage, and behavioral signals** — making top performers immediately obvious and surfacing why others fell short.

COMPARATIVE EVALUATION · TOP CANDIDATES ACROSS PERFORMANCE TIERS — SIDE BY SIDE

CANDIDATE	RESULTS	AI USAGE	NOTABLE SIGNALS	OUTCOME
Sam Rivera TOP Python · 42m	P1–4: 13/13	GPT-5.4 · 3 Contextual	Clean key logic. Strategic AI for edge cases.	✓ STRONG
Taylor Morgan Python · 59m	P1–3: 10/10 P4: 0/3	Opus 4.6 · 5 Descriptive	Never attempted P4. Multiple AI asks on BFS state.	~ MIXED
Morgan Patel Python · 60m	P1: 1/3 P2–4: 0	Haiku 4.5 · 6 Contextual	No visited set. Algorithm loops indefinitely.	⚠ CONCERNS

03
Notable Patterns
Benchmark AI auto-surfaces **what actually separates strong from weak performers** — from decomposition habits to AI usage to edge-case handling — across every interview in your pipeline.

NOTABLE PATTERNS · AUTO-SURFACED — 40 INTERVIEWS

DECOMPOSITION
82% pass rate for small testable functions vs. 34% monolithic

STRATEGIC AI
Top performers: **1–3 prompts**. 5+ correlates with incomplete work

EDGE CASES
15 candidates missed cycle detection. Testing caught it.

STATE MGMT
8 candidates failed P4 BFS state expansion for keys

04
Question Quality
Know **which interview questions are actually doing the work** of differentiating strong from weak performers — and which need to be updated or replaced.

QUESTION QUALITY · CALIBRATION LEVERAGE — MAY 2026

QUESTION	USES	AI FLUENCY	VERDICT
Layer Hierarchy TS, React	12	●●●●○	KEEP
Canvas Cursor TS, React	5	●○○○○	UPDATE
Junior JS JavaScript	2	●○○○○	REPLACE

Compare smarter. Decide faster.
See how Benchmark AI reveals the signal under your candidate scores.

Book a Demo